

Columbia Earthscape:

Overview, Status & Plans

David Millman
Columbia University
dsm@columbia.edu

Topics

- Earthscape
 - business plan, content, technology
 - current status
- Organizational Context – the EPIC center
- Technical Context – architecture
- Plans – new applications, standardization

Earthscape Goals

- to create a fully-integrated, interactive online resource for research and educational materials in Earth Science
- to evaluate the ongoing value and economic viability of providing these services on a subscription-based, cost-recovery model to schools, libraries and individual teachers and students via the Internet

Earthscape Overview

- a publication in the Earth sciences
 - editorial board
 - no corresponding print edition
 - subscription based, self-sustaining
- a collaboration
 - Columbia University Press
 - University library
 - Academic Information Systems
 - partner institutions

Earthscape Business Plan

- startup funding from NSF, SPARC, internal
- subscription income—institutional subscriptions pricing from \$500-\$1500/year; teaching edition at \$250/year
- 500 institutional subscriptions by end of year 3 to become self-sustaining

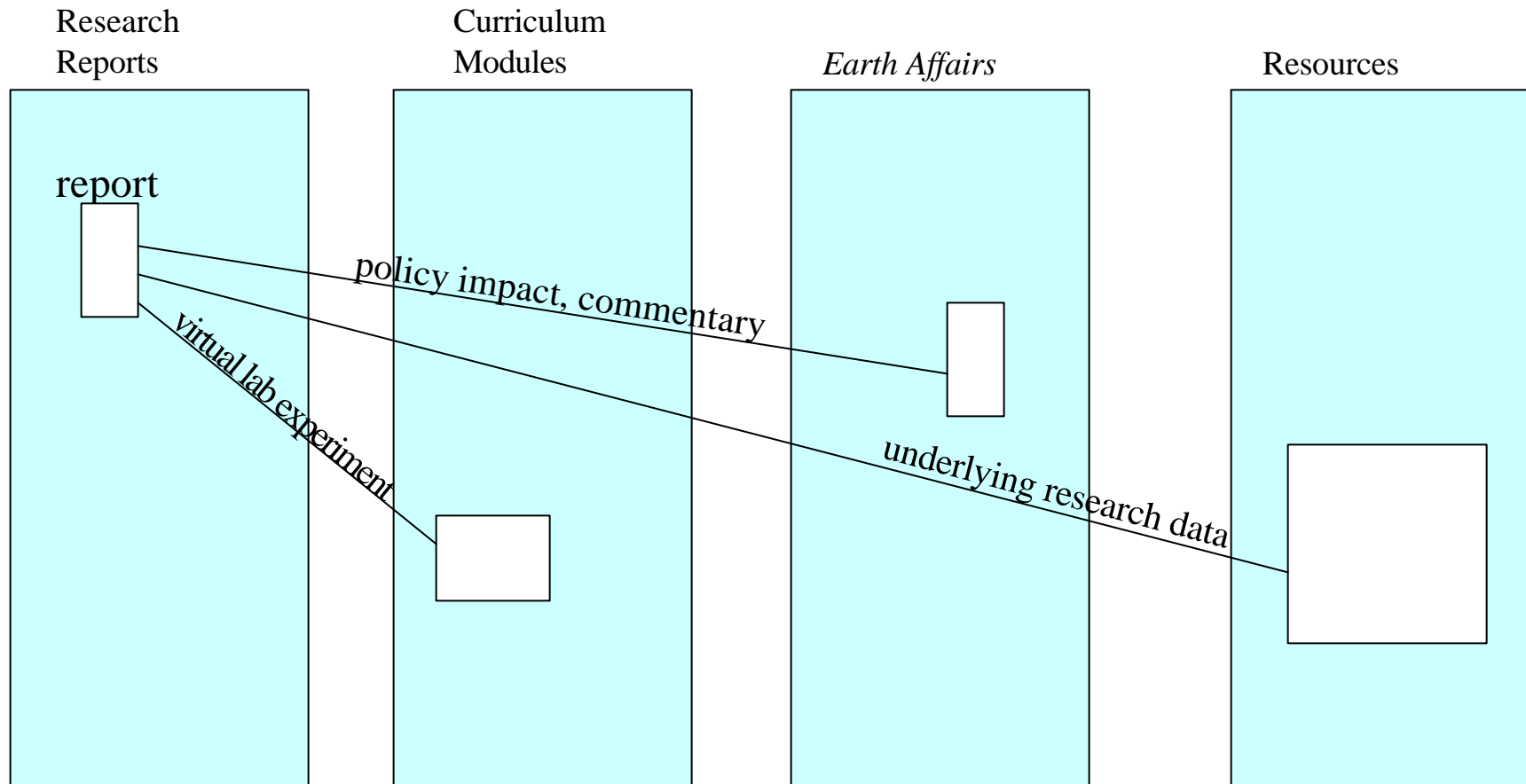
Earthscape Content Plan

1. curricular materials for teaching about the earth from an interdisciplinary and interactive perspective
2. reports of research projects and conference proceedings
3. links to data sets and computer models of ongoing research
4. a magazine, *Earth Affairs*, designed to educate students and the general public

Earthscape Technical Plan

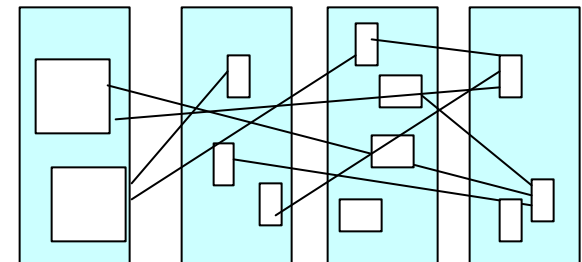
- extensive internal linkage
- citation framework
- multiple classification/term systems
- distributed data/metadata
- distributed security system

Internal Linkage



Linkage technology

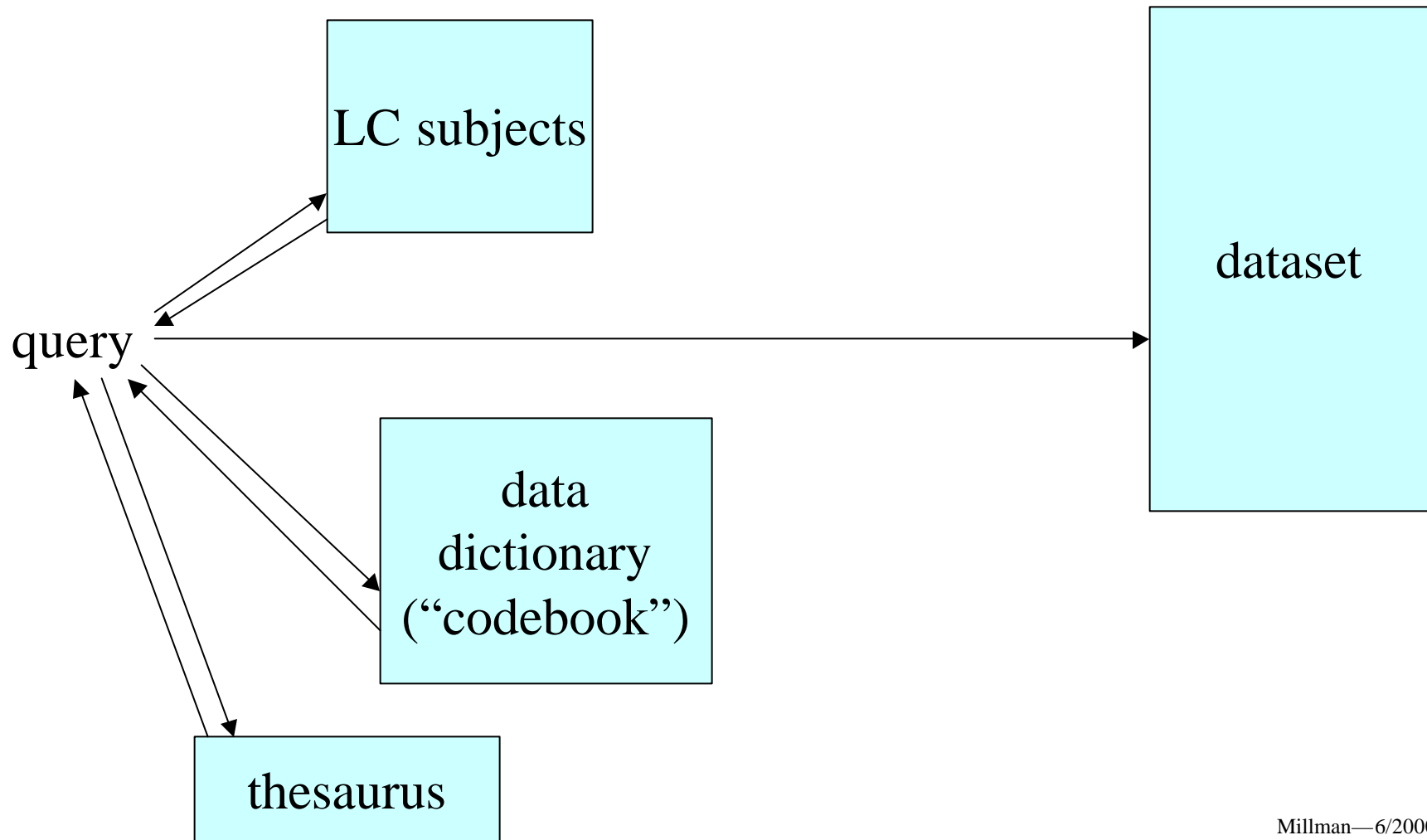
- Naming scheme
- Markup process
- DOI/Handle/PURL
“permanent” identifier



New linkage uses

- Alternate views of site
- Portions as appropriate, per audience
- Citations

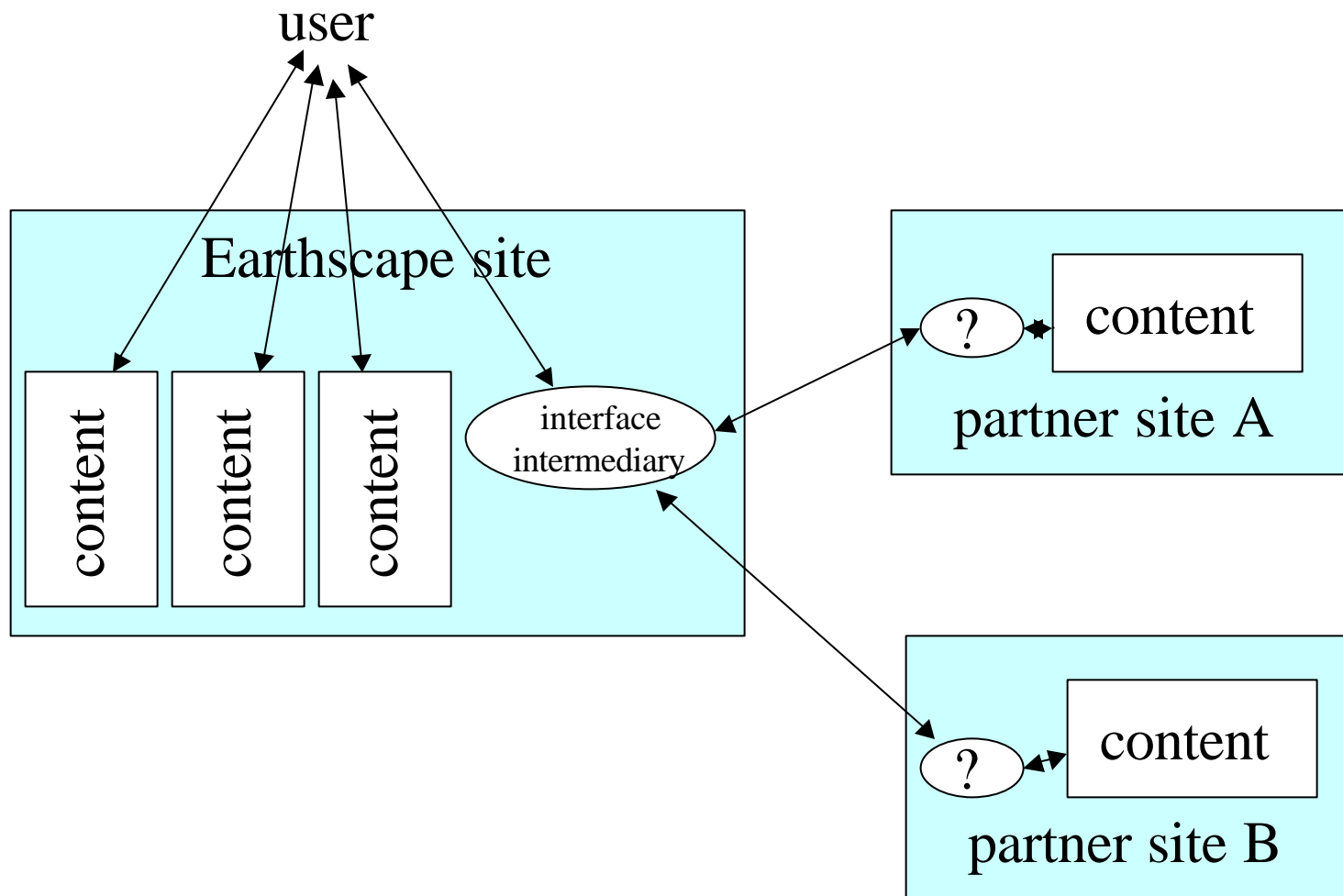
Term Mapping



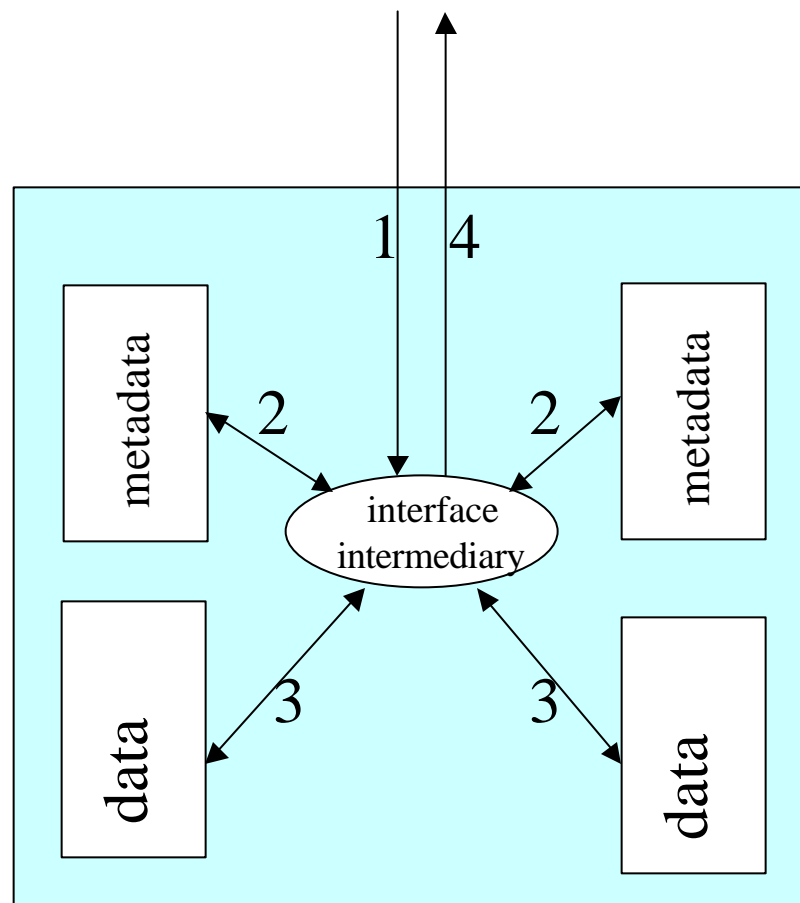
Distributed Architecture

- Deconstructing “interface”
 - kinds of intermediary services
 - resolution of location-independent names
 - kinds of metadata (structural, descriptive, administrative)
 - standards work
- Performance
- Access Management

Distributed architecture

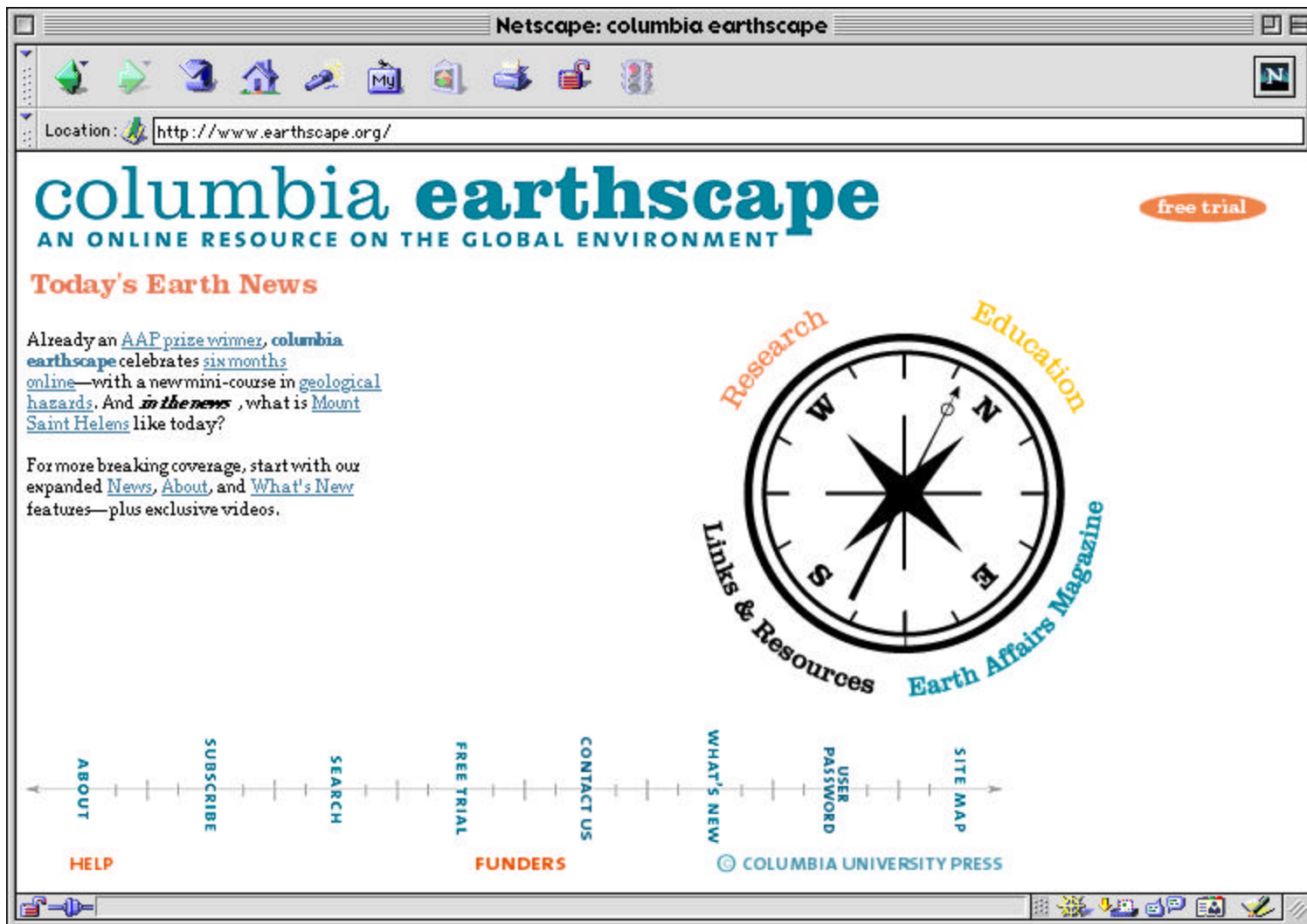


Search across datasets



Status: June, 2000

- Earthscape launched December, 1999
 - 124 curriculum models in Climate Change and Hazards; 50 links to external models
 - 210 research reports, 60 books, 49 journals (abstracts)
 - 2 (+new preview) magazine issues (4-6 papers in each)
 - contains approx 950MB text, still images, & models, plus 30 hours video



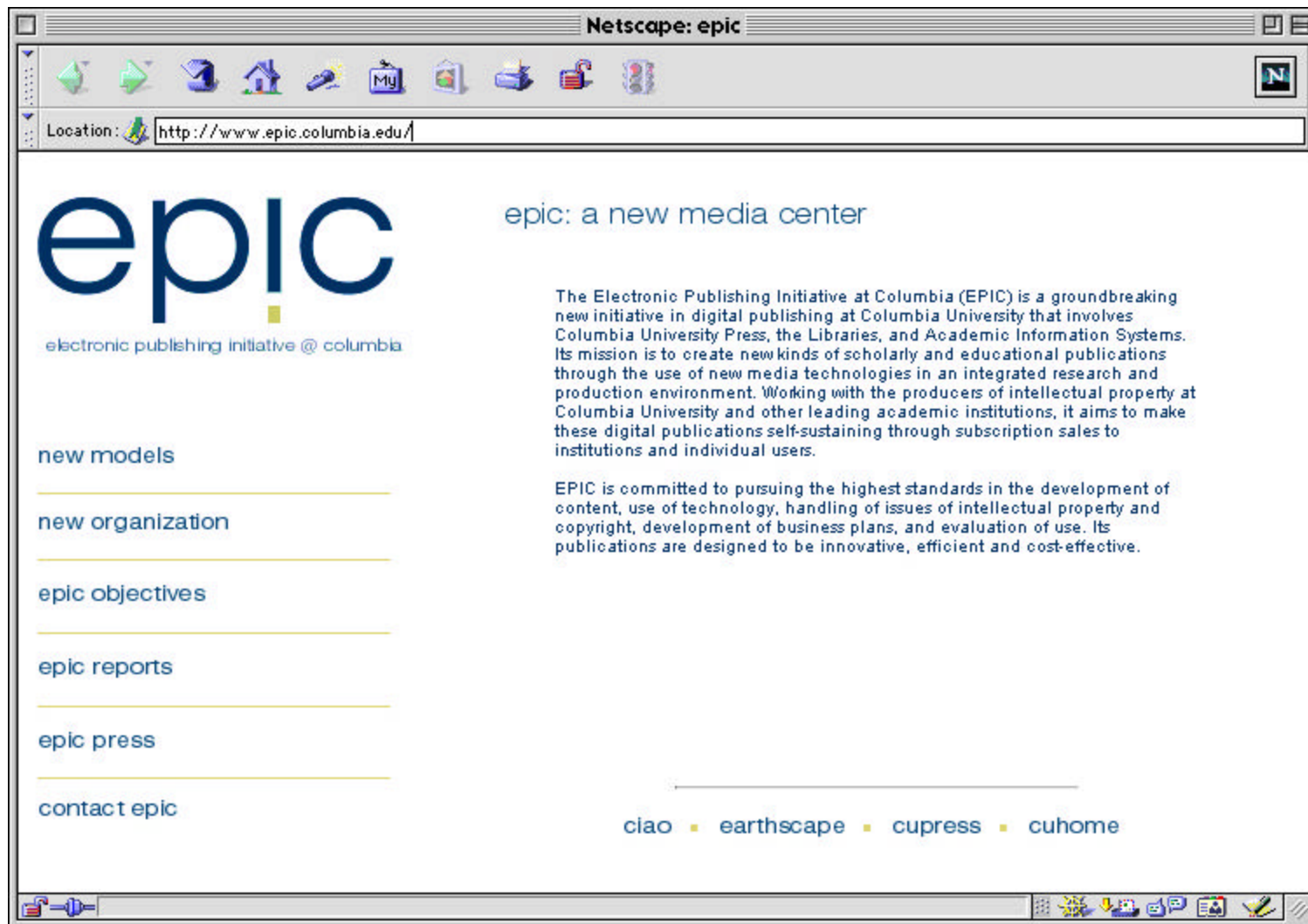
Status . . .

- subscriptions
 - 200 institutions on free trial
 - 55 subscriptions purchased, plus 3 consortia, plus Ohio
- 50 content-contributing partner institutions
 - some outside of the academic community: news organizations (CBS, NY Times), museums
 - creates a rich resource for teaching as well as scholarship

New Organizational Context

- EPIC—Electronic Publishing Initiative at Columbia
- formed by Exec-Vice Provost Sept, 1999
- formalize partnership among Columbia Press, Academic Computing, the Libraries

www.epic.columbia.edu



EPIC—Models in Four Areas

- acquisition, review, and disseminations of online scholarly content
- development of a collaborative organization involving scholars, libraries, technologists, and publishers
- creation of standardized technology systems that can be used across projects
- evaluation of use and effects on research and teaching

EPIC Publications

- Columbia International Affairs Online (CIAO). Launched September, 1997.
- Columbia Earthscape. Launched December, 1999.
- Gutenberg-E. In development. To be launched January, 2001.

EPIC Staffing

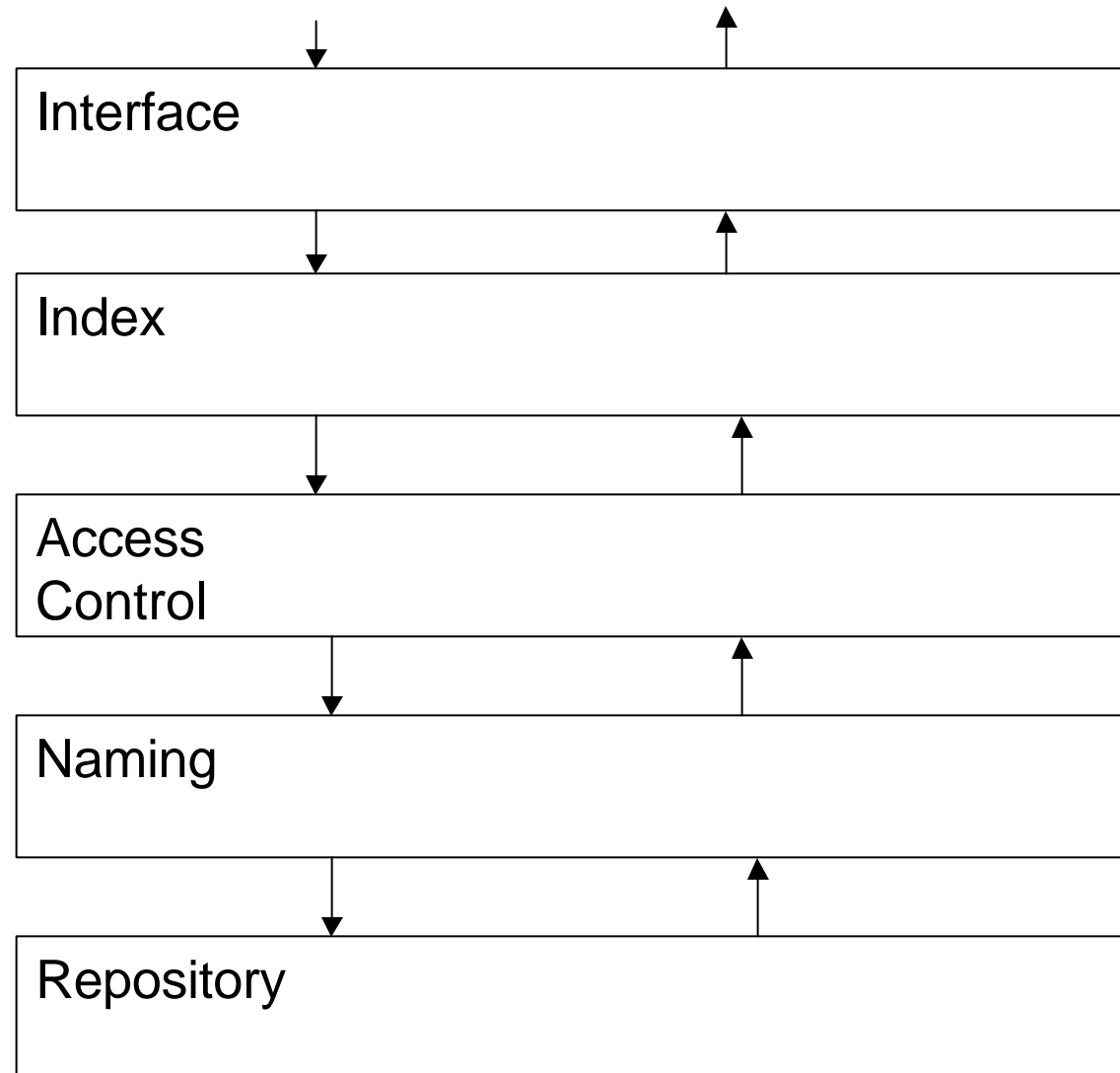
- common
 - management
 - support, marketing
 - evaluation
 - technology development
- per-project
 - editorial
 - contribution to production
 - technology as needed

EPIC Technology

Uses Columbia Digital Library Architecture

- independent layers
- distributed
- modular combinations of facilities
- multi-purpose uses of content

Architecture Overview



1. Repository

- storage of bits
- content neutral
- priorities
 - preservation
 - integrity
- e.g.,
 - backup, redundancy, signed objects

2. Naming

Interface

Index

Access
Control

Naming

Repository

2. Naming

- persistent identifiers
- e.g., DOI, PURL, handles
- for consistent access
- for citations
- issues
 - administration
 - semantics
 - granularity
 - hierarchy/inheritance

3. Access Control

Interface

Index

Access
Control

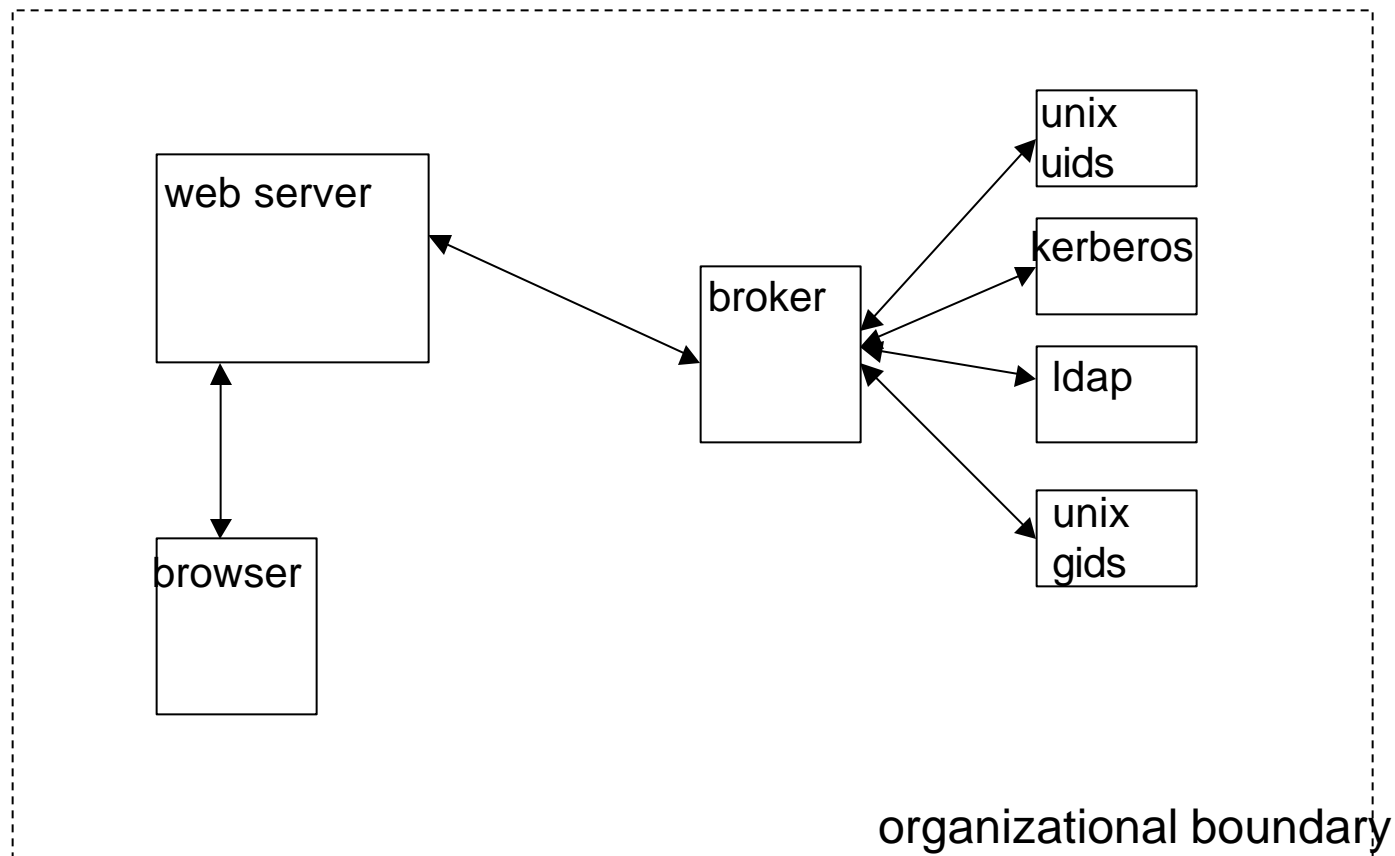
Naming

Repository

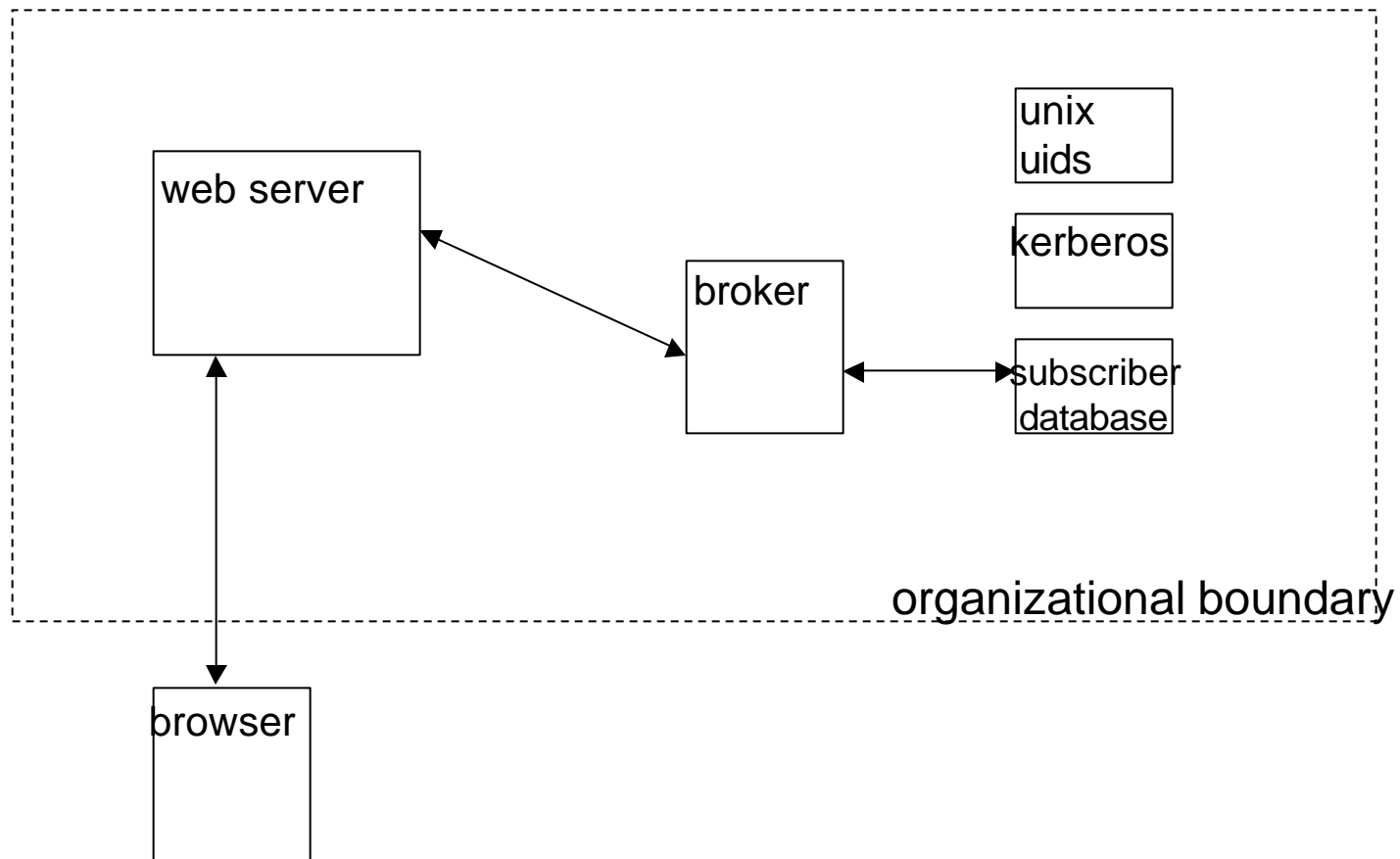
3. Access Control

- higher-ed is both consumer and provider
- “broker” architecture—among authentication (identity) and authorization (capability) systems (dlib.org, Sept 1998)
- certificate architecture—for both auth & auth, highly distributed (dlib.org, Nov 1999)
- specialized local subscription systems
- *toward no IP-based security*

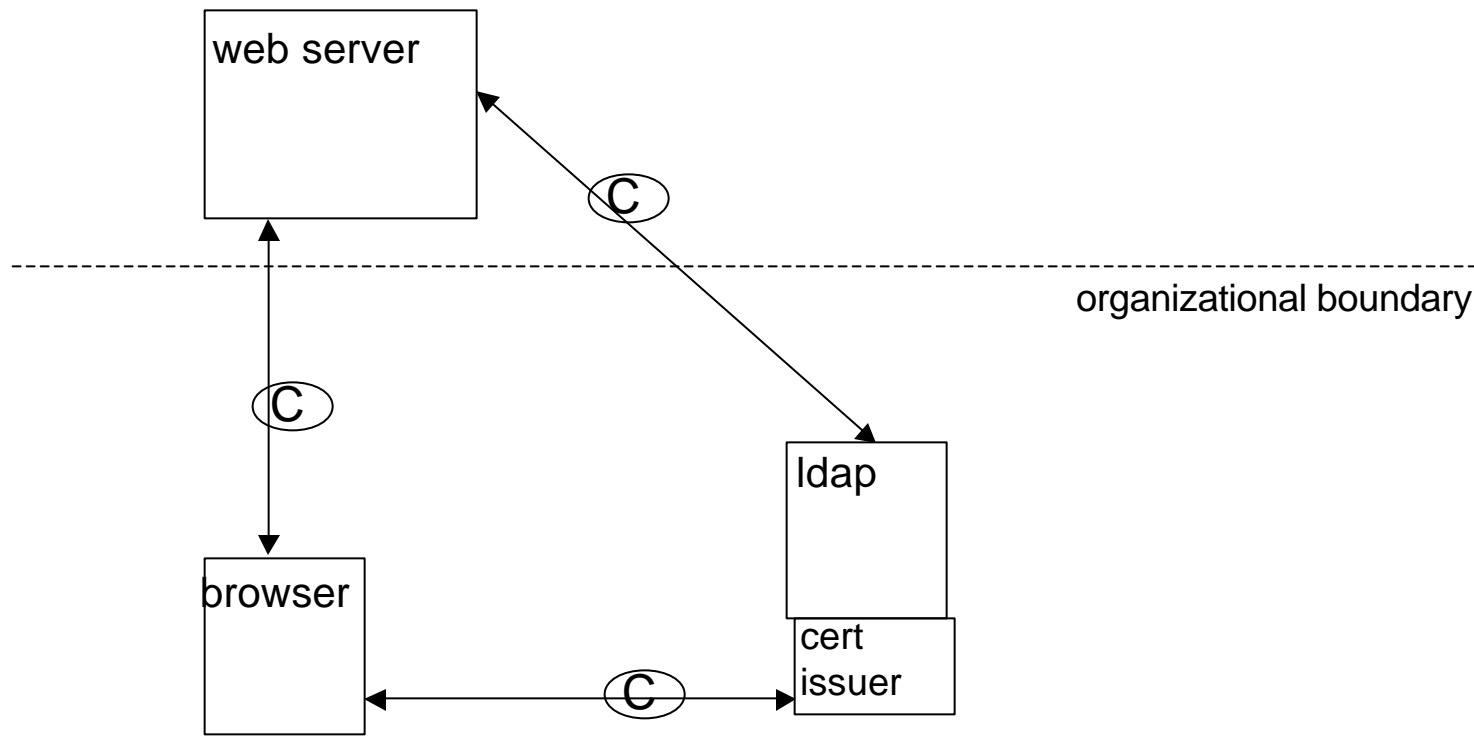
Access Control as campus infrastructure



Access Control as publisher



Access Control as either customer or publisher



Access Control

- technology pilot using digital certificates, with DLF, UC System, U of Minn, OCLC, JSTOR
- policy pilot with CREN CA and enterprise stakeholders (counsel, controller, secretary)
- related work: Internet-2 middleware

4. Index

Interface

Index

Access
Control

Naming

Repository

4. Index

- classification, retrieval, collation
- full-text retrieval
 - performance vs flexibility
- structured text retrieval
- specialized indexes
- master metadata file (MMF)
 - traditional bibliographic/descriptive
 - structural (pages, hierarchies)

5. Interface

Interface

Index

Access
Control

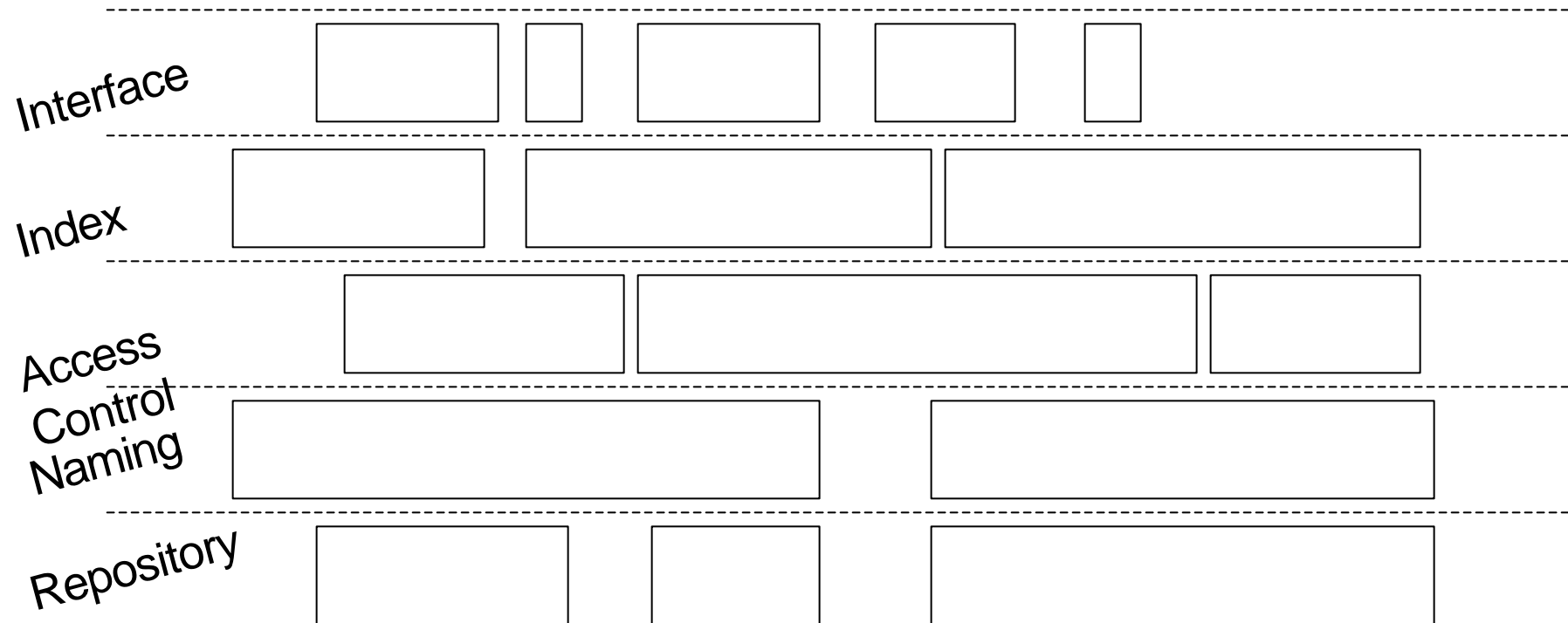
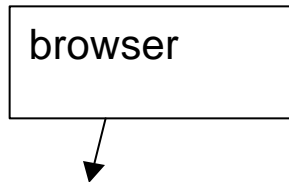
Naming

Repository

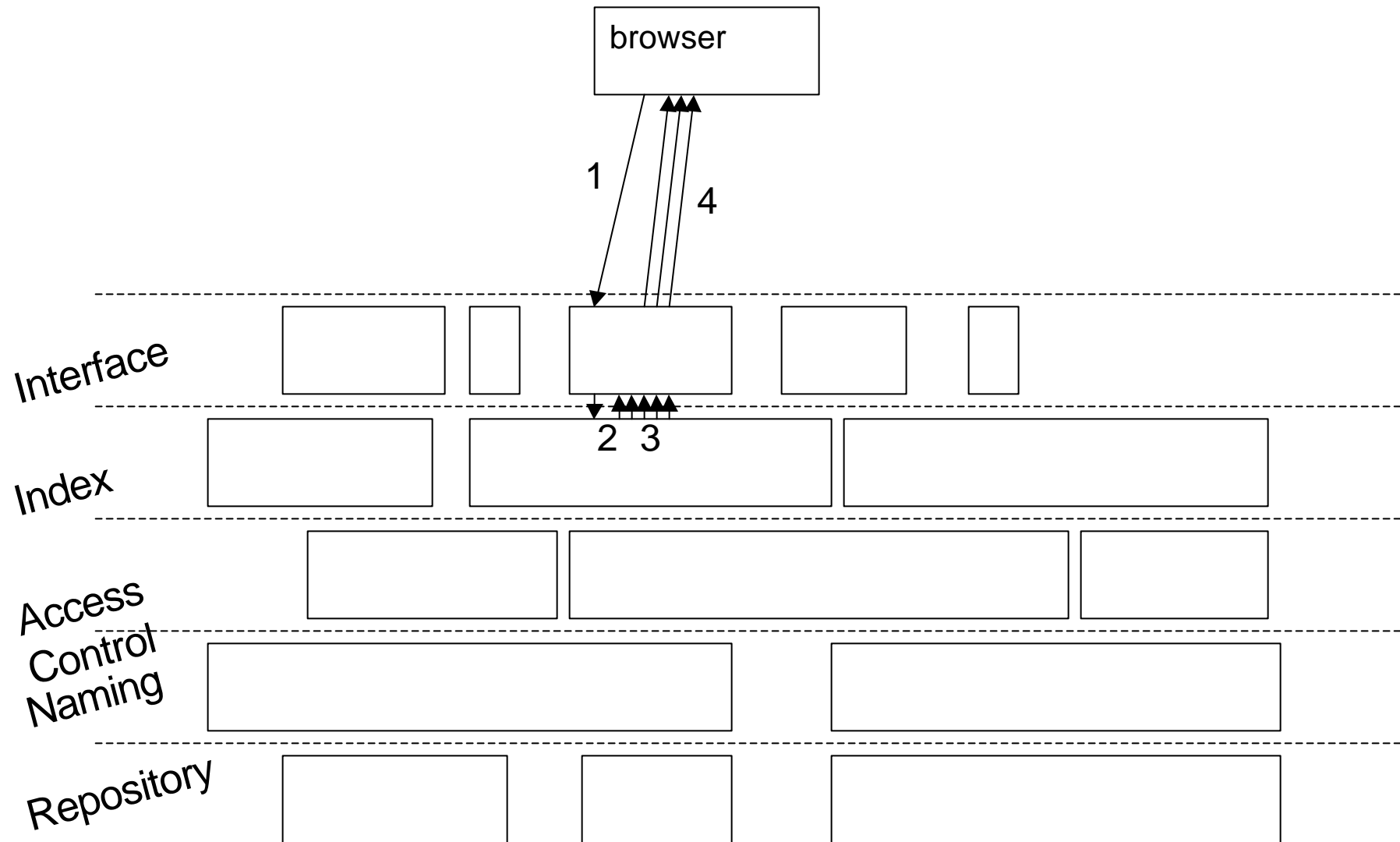
5. Interface

- experience
- conception, description, funding
- vs framework enabling multiple experiences
- content re-use
- integrated content, not links to whole web pages

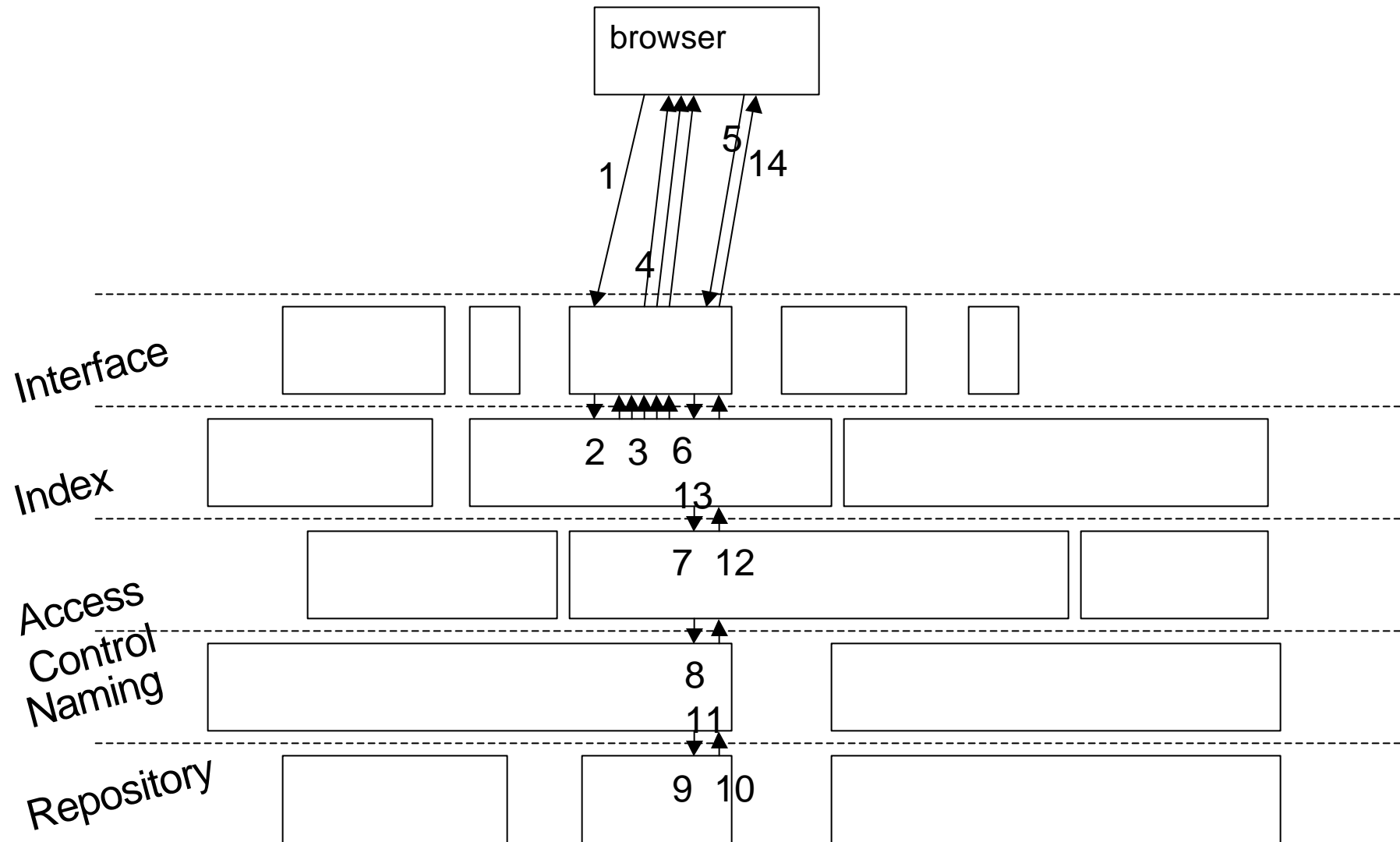
DL Applications



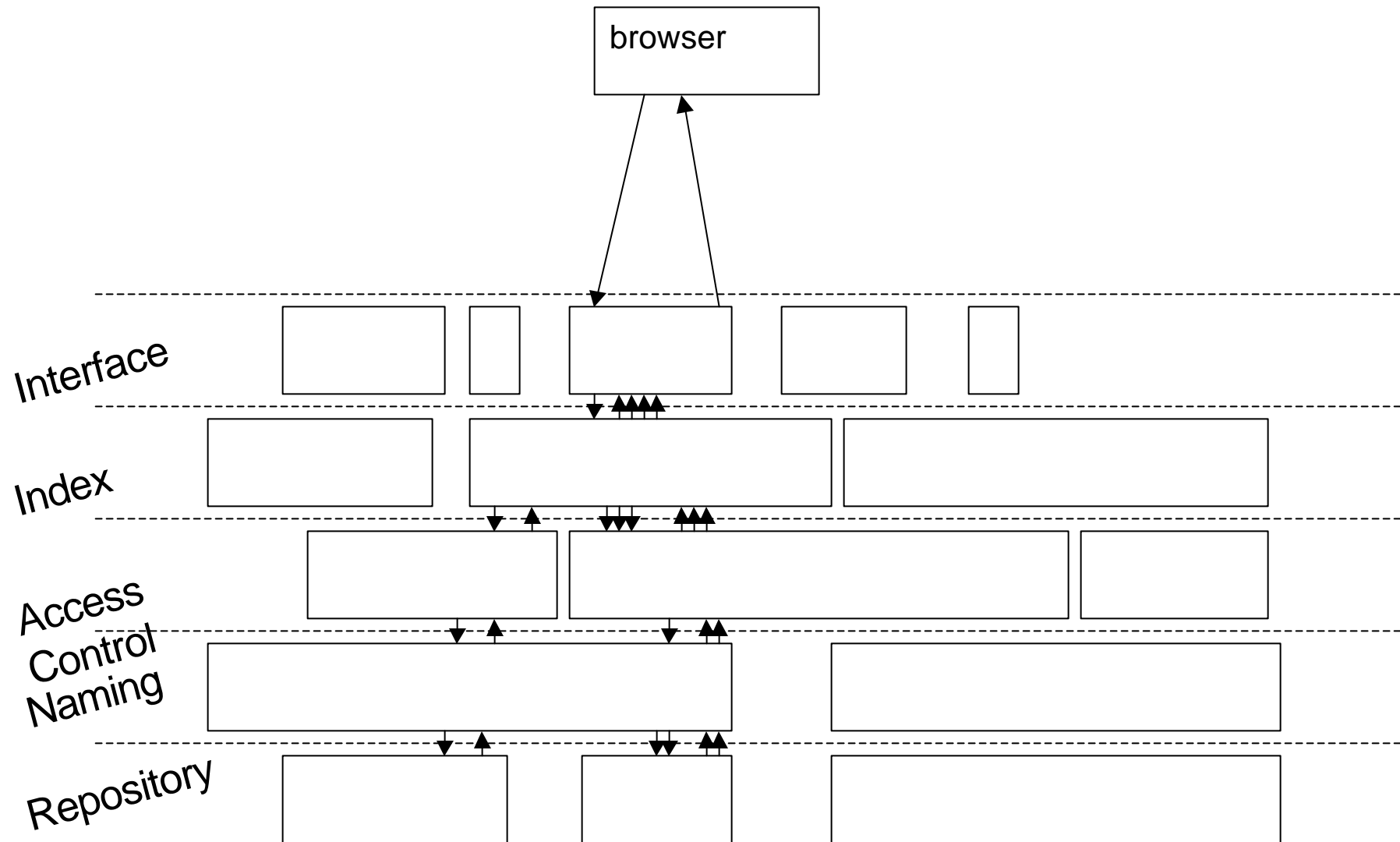
DL Applications



DL Applications



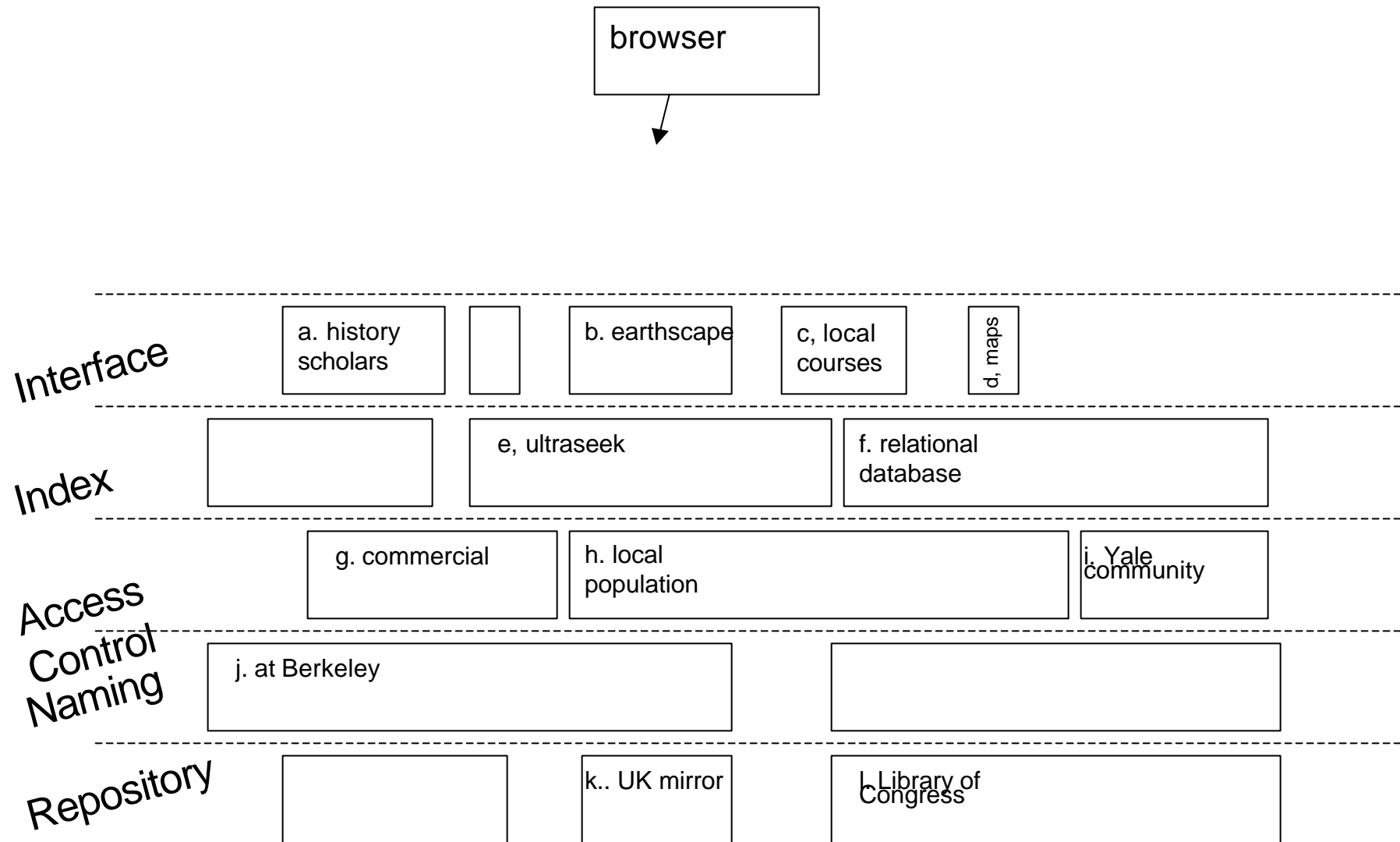
Applications--Interoperable



Scalability

- interoperable layers may be
 - remote
 - have different policies
 - use different technologies
 - have different audiences

DL Applications



Further Work

- extend community pilots in scalable access management
- pursue collaborations in descriptive metadata interoperability (e.g., DLF, Open Archives)
- pursue collaborations in structural metadata interoperability (e.g., object types, harmonization of numerical dataset queries)
- follow preservation and archiving guidelines as they may emerge

Further Work

- develop “depository” with formal properties
 - rights assignment
 - conditions of use
 - mechanism for usage reporting
 - mechanism for author/publisher credit/payment
 - document authenticity

Depository Funding Models

- institutional subscription model
- free use through support from funding agencies
- sponsorship by commercial organization
- access fees for some and free access to others depending on ability to pay

Standards and Models

- technology
- staffing
 - technology, design, production
 - editorial
- business
 - licensing, intellectual property management
 - sustainability
- partnerships
- archiving and preservation

fin